ValuesRAG: Enhancing Cultural Alignment Through Retrieval-Augmented Contextual Learning

Wonduk Seo[1*] **Zonghao Yuan**[2*] **Yi Bu**[1,3†]
[1]Peking University [2]Tsinghua University
[3]Peking University Chongqing Research Institute of Big Data
[1]{seowonduk, buyi}@pku.edu.cn, [2]yzh23@mails.tsinghua.edu.cn

Abstract

Cultural values alignment in Large Language Models (LLMs) is a critical challenge due to their tendency to embed Western-centric biases from training data, leading to misrepresentations and fairness issues in cross-cultural contexts. Recent approaches, such as role-assignment and few-shot learning, often struggle with reliable cultural alignment as they heavily rely on pre-trained knowledge, lack scalability, and fail to capture nuanced cultural values effectively. To address these issues, we propose ***ValuesRAG****, a novel and effective framework that applies Retrieval-Augmented Generation (RAG) with In-Context Learning (ICL) to integrate cultural and demographic knowledge dynamically during text generation*. Leveraging the World Values Survey (WVS) dataset, ValuesRAG first generates summaries of values for each individual. Subsequently, we curate several representative regional datasets to serve as test datasets and retrieve relevant summaries of values based on demographic features, followed by a reranking step to select the top-k relevant summaries. ValuesRAG consistently outperforms baseline methods, both in the main experiment and in the ablation study where only the values summary was provided. Notably, ValuesRAG demonstrates an accuracy of 21% improvement over other baseline methods, highlighting its potential to foster culturally aligned AI systems and enhance the inclusivity of AI-driven applications.

1 Introduction

The rapid advancement of Large Language Models (LLMs) has revealed pressing challenges in cultural values alignment [Singh *et al.*, 2024; Kharchenko *et al.*, 2024; Hu *et al.*, 2024]. Predominantly trained on Western data sources [Achiam *et al.*, 2023; Touvron *et al.*, 2023; Jiang *et al.*, 2023], LLMs inherently reflect Western cultural norms and social biases, raising concerns about their applicability in global contexts.

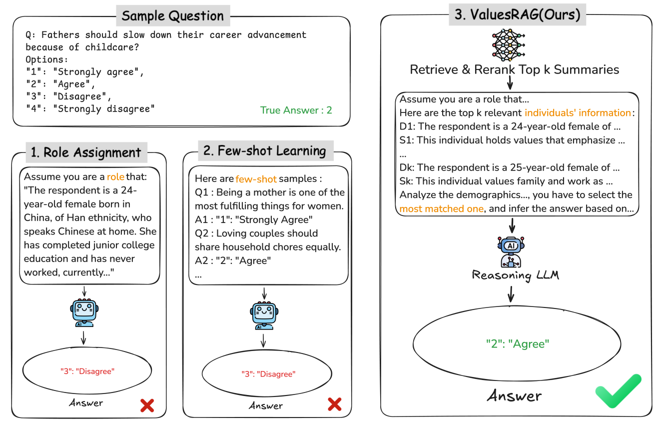

Figure 1: **Overview of different approaches for cultural alignment.** Comparing two baseline methods, namely Role Assignment and Few-Shot Learning, and our proposed ValuesRAG framework.

These biases present significant challenges when deploying LLMs in cross-cultural environments, often resulting in misrepresentations and stereotypical outputs [Gallegos *et al.*, 2024; Xie *et al.*, 2024; Potter *et al.*, 2024; Huang *et al.*, 2024]. Despite ongoing efforts to address these issues, existing strategies often fall short. While some countries have developed localized LLMs, such as China's ERNIE [Sun *et al.*, 2021], ChatGLM [GLM *et al.*, 2024], DeepSeek [Liu *et al.*, 2024a], and South Korea's HyperCLOVA [Yoo *et al.*, 2024], these models also exhibit biases inherited from their respective training datasets. As a result, cultural and social biases embedded in LLMs remain a critical concern, compelling researchers to explore more robust frameworks for cultural alignment [Gallegos *et al.*, 2024; Xie *et al.*, 2024; Potter *et al.*, 2024].

Recent studies have proposed several approaches, such as *role-assignment* approaches [Tao *et al.*, 2024] and *few-shot learning* techniques [Choenni and Shutova, 2024], to mitigate these cultural biases. However, these methods still face several challenges: (1) Role-assignment approaches, relying solely on the model's pre-trained knowledge, provide pre-defined demographic information but fail to incorporate explicit values alignment text, which subsequently introduces stereotypes and biases rooted in Western-centric train-

*Wonduk Seo and Zonghao Yuan contributed equally to this work.
†Yi Bu is the corresponding author for this work.

ing data; (2) While offering example-based guidance, few-shot learning methods struggle to comprehensively capture the complex cultural values due to the limited correlation between different values dimensions, thus remain ineffective on values-related tasks that differ significantly from the examples; (3) In addition, these methods can only align with the values of a single individual, and singular values cannot represent the universal values of individuals with similar characteristics.

To address these challenges, we propose ***ValuesRAG**, a novel framework that utilizes Retrieval-Augmented Generation (RAG) and In-Context Learning (ICL) to dynamically incorporate cultural knowledge during text generation* (see Figure 1). Our framework leverages the World Values Survey (WVS) dataset [Haerpfer *et al.*, 2022], a globally recognized and comprehensive dataset that explores values across countries using rigorous social science methodologies. Specifically, we first generate summaries for each topic, followed by generating individuals' summaries of values and demographic profiles in parallel. After constructing the knowledge base, we retrieve the top 100 relevant summaries based on demographic features, followed by a reranking step to ensure the most relevant top-k summaries are selected. Finally, we utilize a reasoning LLM that filters the most relevant demographic profiles and applies reasoning grounded in the retrieved values to generate the final answer to the question.

We evaluate the performance of ValuesRAG by comparing it against several baseline approaches, including: (1) *zero-shot inference*, (2) *role-assignment-only method* [Tao *et al.*, 2024], (3) *few-shot learning* [Choenni and Shutova, 2024], and (4) *a hybrid method combining (1) and (2)*. To ensure a comprehensive evaluation, we curated diverse regional survey QA datasets which are designed to capture values-related question-answer pairs. Extensive experimental results show significant improvements in cultural and contextual understanding, demonstrating that ValuesRAG outperforms the baselines. Unlike previous methods that heavily depend on pre-trained knowledge or limited demonstrations, ValuesRAG dynamically retrieves and integrates multiple similar individual values summaries based on demographic features, enabling richer value representations and more context-aware responses compared to approaches relying on a single predefined prompt or role.

In addition, ablation studies on using only value-augmented generation and on varying the number of retrieved summaries confirm ValuesRAG's robust performance under different configurations. ValuesRAG surpasses the baselines through purely values-based generation. Meanwhile, adjusting the number of retrieved documents shows that moderate retrievals can effectively balance diversity and relevance, also maintaining high accuracy across multiple benchmarks.

These findings highlight ValuesRAG's potential to foster inclusive AI systems, enhancing the reliability and fairness of AI-driven applications. Our study demonstrates ValuesRAG's robust capabilities on a global scale, also suggesting its applicability in aligning the values of diverse groups within a single country. ValuesRAG provides a cost-efficient tool for public policymakers and scientists from various disciplines to refine social simulations, enabling more precise predictions of policy outcomes [Li *et al.*, 2024]. This, in turn, facilitates the creation of fairer and more effective policies. Moreover, NGOs can leverage ValuesRAG to develop LLMs that reflect specific value orientations while maintaining strong alignment with users' values, thereby increasing their persuasive impact. This approach benefits the promotion and spread of values that contribute to the planet's sustainable development and the long-term well-being of human society.

2 Related Work

2.1 Evaluation of LLMs' Cultural Bias

Pre-trained models are facing growing criticism for their inherent social biases, with cultural bias emerging as a particularly nuanced and pervasive issue [Tao *et al.*, 2024]. Unlike the more obvious safety concerns and social discrimination [Liu *et al.*, 2024b] embedded in language models, cultural bias manifests in subtler ways, often reflecting the dominant cultural perspectives present in training data. Studies have shown that LLMs often exhibit cultural biases aligned with the values of developed countries, resulting in the underrepresentation of perspectives from less developed regions [Manvi *et al.*, 2024; Durmus *et al.*, 2024]. This imbalance not only perpetuates existing cultural hierarchies but also limits the global applicability of these models [Manvi *et al.*, 2024]. Various benchmarks and evaluation methods have been proposed to assess the cultural biases of pre-trained models [Gallegos *et al.*, 2024]. For example, Webster et al. [2021] developed probability-based metrics to evaluate gender bias embedded in pre-trained models, while Caliskan et al. [2017] pioneered the use of word embeddings as quantitative measures of bias. More recently, Karinshak et al. [2024] introduced *LLM-GLOBE*, a benchmark where LLMs generate both quantitative and open-ended answers to values assessment questions, with subsequent evaluation using the LLM-as-a-Jury Protocol. These evaluation methods collectively highlight the complex nature of cultural bias in LLMs and the need for multifaceted assessment approaches.

2.2 Mitigation of LLMs' Cultural Bias

Methods like RLHF [Shen *et al.*, 2023; Ji *et al.*, 2024] are very commonly used in LLM values alignment, but this type of single values alignment method is not that suitable for the mitigation of cultural bias, because the alignment goal of cultural bias is diverse and dynamic [Huang *et al.*, 2024], as there are hundreds of countries and cultures in the earth. Addressing cultural biases has become a critical area of research, with various strategies being proposed to enhance cultural sensitivity in LLMs. For instance, Tao et al. [2024] adopted national and cultural role assignments to adjust the cultural values of LLMs, while Masoud et al. [2024] developed a soft prompt tuning approach to mitigate bias. Also, Choenni and Shutova [2024] employed few-shot in-context learning to align cultural behaviors, demonstrating promising results in specific contexts. However, these approaches face significant limitations in fully capturing the complexity of cultural alignment. Tao et al.'s technique mainly depends on national

Category	Dataset	Abbreviation	Region	Year	N	VQ
Retrieval Corpus	World Values Survey	WVS	Global	2017–2022	97.2k	259
Test Datasets	European Values Study	EVS	Europe	2017	59.4k	211
	The General Social Survey	GSS	North America	2021–2022	8.2k	44
	Chinese General Social Survey	CGSS	East Asia	2021	8.1k	58
	India Survey Dataset	ISD	South Asia	2019–2020	30.0k	33
	AmericasBarometer	LAPOP	Latin America	2021	59.1k	48
	Afrobarometer	Afrobarometer	Africa	2022	48.1k	144

Table 1: **Overview of the datasets utilized in our study.** The *Retrieval Corpus* (WVS) includes global data collected between 2017 and 2022, providing the basis for generating cultural summaries of values and validation for our method. The *Test Datasets* consist of six region-specific surveys, each capturing socio-cultural information from distinct geographic areas and time frames. N represents sample size in thousands (k). VQ represents the number of values-related questions.

and cultural roles without explicitly integrating values assignments, causing an overreliance on latent internal representations. Meanwhile, Choenni and Shutova's few-shot learning approach similarly falls short of modeling cultural alignment in all its complexity. We therefore use these methods as baselines to benchmark our proposed approach.

3 Datasets

In this section, we first introduce the World Values Survey (WVS) as our retrieval corpus (Section 3.1), highlighting its extensive coverage, global representativeness, and relevance for values-related studies. Subsequently, we describe six regional test datasets (Section 3.2), which are carefully selected to ensure geographic, cultural, and demographic diversity.

3.1 Retrieval Corpus

VS [1] [Haerpfer *et al.*, 2022] is a globally recognized dataset that investigates human beliefs, values, and cultural norms through structured surveys conducted across multiple countries. WVS is selected as our retrieval corpus especially due to its numerous advantages:

1. Broad recognition and inclusiveness: WVS is widely recognized and frequently used by governments, social scientists, and major international organizations in comparative values studies. It currently covers 120 countries, representing 94.5% of the global population, ensuring broad geographic and cultural representation.

2. Expert-designed and accessible: The dataset is meticulously designed by leading domain experts to conduct comprehensive surveys of values, ensuring reliability, rigor, and relevance. It is publicly accessible, enabling reproducibility and transparency in research.

3. Effective structure and large scale: WVS has well-organized and comprehensive demographic questions, making it effective for retrieval tasks. Its large sample size (97,221 respondents) is also suitable for RAG tasks.

Since values evolve gradually over time, VS is conducted in waves, with each wave occurring every five years. For our study, we utilize the most recent wave, spanning from 2017 to 2022.

[1]https://www.worldvaluessurvey.org/wvs.jsp

The WVS codebook includes over 600 indicators, with 259 values-related and 31 demographic-related questions. The value questions span 13 topics, such as social trust, post-materialism, and political interest. We randomly select 20% (52 questions) per topic for validation and use the remaining 80% (207 questions) for summary generation. The 31 demographic features, including country, sex, age, education, social class, and employment status, are used to generate demographic summaries for retrieval tasks.

3.2 Test Datasets

We select six regional surveys[2] to serve as test datasets based on the following criteria:

1. Demographic and values coverage: The datasets provide demographic features closely aligned with WVS's questions, along with sufficient values-related features to enable meaningful comparisons and analyses.

2. Temporal proximity: The datasets exhibit close temporal proximity to WVS Wave 7 (2017–2022), thereby allowing aligned comparisons and ensuring thorough consistency across diverse global evaluations.

The regions in our test datasets are meticulously chosen to encompass a wide range of geographic, cultural, and demographic diversity, ensuring that the data accurately reflects the majority of the global population. All of them are publicly accessible and are statistically representative at national or regional levels, which guarantees their reliability and validity.

To be specific, we select the European Values Study [EVS, 2022] as the representative dataset for Europe, as it is the largest values survey in the region. For the United States, we select the General Social Survey [Davern *et al.*, 2024], which is the most comprehensive social survey in the country. The Chinese General Social Survey [Bian and Li, 2012] serves as the representative dataset for China due to its comprehensive sampling methodology and scientific rigor. For India, where national survey data were largely inaccessible during our study, we use Pew Research Center's survey data [Sahgal and Evans, 2021] to represent the Indian population. The AmericasBarometer [Lab, 2021], conducted by the LAPOP

[2]Detailed dataset description are provided in Appendix A.

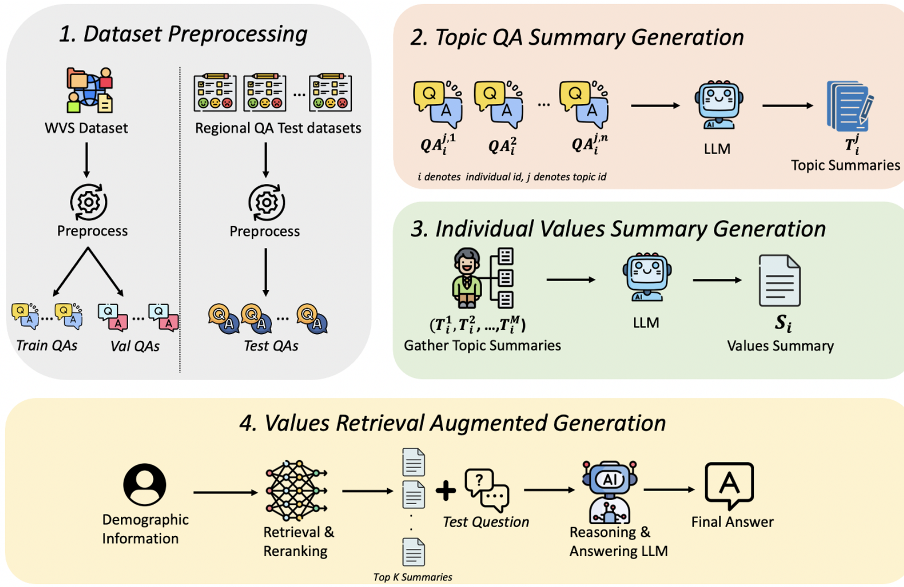

Figure 2: **Overview of the proposed ValuesRAG framework for cultural alignment.** Consisting four components: (1) Dataset Preprocessing, (2) Topic QA Summary Generation, (3) Individual Values Summary Generation, and (4) Values Retrieval Augmented Generation.

Lab, is selected to represent Latin American countries, as it covers 32 countries across the region. Finally, Afrobarometer [Afrobarometer, 2023] is chosen as the representative dataset for Africa. A detailed summary of these datasets is presented in Table 1.

4 Methodology

In this section, we present the *ValuesRAG* which is specifically designed to address cultural biases and enhance contextual alignment in LLM-driven scenarios through a Retrieval-Augmented Generation (RAG) approach. *ValuesRAG* consists of three key components: *Values and Demographic Summary Generation*, which extracts and summarizes cultural values and demographic information from large-scale datasets (Section 4.1); *Values-Augmented Generation*, which incorporates these summaries into the generative process to align responses with the cultural context (Section 4.2); and *Retrieval-based Values Alignment*, which dynamically assigns relevant individual values to queries based on demographic profiles (Section 4.3).[3] An overview of the *ValuesRAG* framework is provided in Figure 2.

4.1 Values and Demographic Summary Generation

systematically generate concise summaries of values and demographics for each individual, we process the dataset in three stages. First, the dataset is stratified by topics and split into train and validation sets, ensuring that the distribution of each topic is preserved (described in Section 3.1). In parallel, topic-based summaries and demographic summaries are generated separately. For topic-based summaries, values-related QA sets are used to produce summaries for each topic, while demographic summaries are generated using demographic-related QA sets:

$$\begin{aligned} T_i^j &= f_{\text{gen}}(\text{QA}_i^{j,1}, \text{QA}_i^{j,2}, \ldots, \text{QA}_i^{j,N_j}), \\ D_i &= f_{\text{gen}}(\text{QA}_i^{\text{demo},1}, \text{QA}_i^{\text{demo},2}, \ldots, \text{QA}_i^{\text{demo},K}). \end{aligned} \tag{1}$$

where f_{gen} denotes the generative model, T_i^j is the summary for topic j of individual i, based on N_j values-related QA pairs, and D_i represents the demographic summary derived from K demographic-related QA pairs. Finally, individual summaries are constructed by combining all topic summaries:

$$S_i = f_{\text{gen}}(T_i^1, T_i^2, \ldots, T_i^M), \tag{2}$$

with M denoting the total number of topics. The result, denoted as S_i, forms a comprehensive values summary for individual i. These generate summaries serve as structured references for retrieval in later stages and are also used to augment the validation set for evaluation.

4.2 Values Augmented Generation

Once the comprehensive summaries for each individual are generated in the previous step, we construct an augmented generation process for evaluating on the validation question-answer data. For each validation question, we concatenate the corresponding individual's values summary with the question itself, forming a context-rich input for the LLM:

$$C_i = \text{concat}\left(S_i, \text{Q}_i^{\text{val},k}\right) \tag{3}$$

where C_i represents the combined context, S_i is the values summary for individual i, and $\text{Q}_i^{\text{val},k}$ is the k-th validation question. Next, we concatenate C_i with the demographic summary D_i to further enhance the context, enabling the generation of responses based on both values and demographic information:

$$A_i = f_{\text{gen}}\left(C_i, D_i\right) \tag{4}$$

[3]We provide a case study of retrieve in Appendix C.

Here, A_i represents the answer generated by the function f, and (C_i, D_i) embeds the augmented context C_i and demographic information D_i into a structured input format. Additionally, we utilize chain-of-thought prompting to enhance reasoning and emulate the behavior of the corresponding individual, ensuring responses that are contextually aligned with the values captured in the summaries and demographic characteristics.

4.3 Retrieval-based Values Alignment

To dynamically assign relevant values to test individuals, we leverage demographic information as documents for retrieval. The demographic data from both the train (retrieval corpus) and test datasets are preprocessed into a structured context format, as described earlier, and embeddings are generated for each demographic context using a representation model. We first retrieve the top-100 most similar summaries of values for each test individual by computing the cosine similarity between the embeddings of the test and train demographics:

$$\mathrm{Sim}(E_{\text{test}}, E_{\text{train}}) = \frac{E_{\text{test}} \cdot E_{\text{train}}}{\|E_{\text{test}}\| \|E_{\text{train}}\|} \tag{5}$$

Specifically, E_{test} and E_{train} represent the embeddings of the test and training (retrieval corpus) demographic contexts, respectively, and $\mathrm{Sim}(\cdot, \cdot)$ denotes the cosine similarity score. The top-100 embeddings with the highest similarity scores are initially selected as candidates. Subsequently, we apply a reranking step to refine the selection and identify the most relevant summaries among the retrieved candidates. The reranking process evaluates the semantic relevance of the initial candidates based on a scoring function:

$$R'_k = f_{\text{rerank}}\left(E_{\text{test}}, E_{C_j}\right),\ C_j \in \{C_1, C_2, \ldots, C_{100}\},\ k \leq 100 \tag{6}$$

where R'_k represents the reranked summary for the k-th candidate selected from the final top-k results, f_{rerank} is the reranking function, and E_{C_j} denotes the embedding of the j-th candidate initially retrieved among the 100 documents. The reranked top-k summaries are then incorporated into the prompts, enriching the contextual alignment of the generated responses. In detail, for each test individual, the retrieved and reranked summaries are combined into the final prompt, and the answer is subsequently generated using the function f_{gen}:

$$\begin{aligned} P_{\text{test}} &= (D_{\text{test}}, R'_1, R'_2, \ldots, R'_K, Q_{\text{test}}), \\ A_{\text{test}} &= f_{\text{gen}}(P_{\text{test}}). \end{aligned} \tag{7}$$

Here, P_{test} is the final prompt, D_{test} is the demographic information of the test individual, $\{R'_1, R'_2, \ldots, R'_K\}$ represents the top-k reranked summaries, and Q_{test} is the test question. A_{test} denotes the generated answer for the test question, and f_{gen} represents the generation function.

This retrieval-based approach, followed by reranking, enhances reasoning by explicitly guiding the LLM to critically evaluate which retrieved values best align with the test individual's demographic characteristics. The final prompts are then used to generate answers following the chain-of-thought prompting strategy, ensuring that the responses are contextually coherent and culturally aligned with the test individual's profile. For the comprehensive implementation of the *ValuesRAG*, we provide the **Algorithm 1**, which systematically outlines the processes of values and demographic summary generation, values-augmented generation, and retrieval-based values alignment, as shown below:

Algorithm 1 *Values Generation and Retrieval Process*

Require: Dataset $\mathcal{D}$ with topics and demographic QA pairs, Generative Model f_{gen}, Embedding Model f_{embed}, Reranking Model f_{rerank}, Retrieval Top-K

// Values and Demographic Summary Generation
for each individual $i \in \mathcal{D}$ **do**
  Generate topic-based values summaries T_i^j for each topic j
  Generate demographic summaries D_i
  Combine T_i^j into comprehensive values summary S_i
end for
// Values Augmented Generation
for each validation question $Q_i^{\text{val},k}$ of individual i **do**
  Construct context $C_i = \mathrm{concat}(S_i, Q_i^{\text{val},k})$
  Augment context with demographic summary D_i
  Generate answer $A_i = f_{\text{gen}}(C_i, D_i)$
end for
// Retrieval-based Values Alignment
for each test individual i **do**
  Compute embeddings $E_{\text{test}} = f_{\text{embed}}(D_{\text{test}})$
  Retrieve top-100 values summaries by similarity: $\mathrm{Sim}(E_{\text{test}}, E_{\text{train}})$
  Rerank top-K summaries: $R'_k = f_{\text{rerank}}(E_{\text{test}}, E_{C_j})$
  Final prompt $P_{\text{test}} = (D_{\text{test}}, R'_1, \ldots, R'_K, Q_{\text{test}})$
  Generate answer $A_{\text{test}} = f_{\text{gen}}(P_{\text{test}})$
end for

5 Experiments

5.1 Setup

Models Used. We utilize *GPT-4o-mini* [Achiam *et al.*, 2023] and *Gemini-1.5-Flash* [Team *et al.*, 2024] for our generation tasks, which are accessed via APIs. We set the temperature parameter of these models to 0.7 to achieve a balance between coherence and creativity. For the retrieval task, we employed the *E5 (base)* model [Wang *et al.*, 2022][4], which generates embeddings and retrieves the top 100 most relevant summaries of values based on cosine similarity. Additionally, we utilize the *GTE-multilingual-reranker-base model* [Zhang *et al.*, 2024][5] for reranking. This reranking model refines the selection process and ensures that the top-k most relevant summaries are selected from the initial top-100 retrievals.

Baseline Methods and Implementation. Our baseline methods include: (1) *Zero-shot inference*, (2) *the role-assignment-only approach* [Tao *et al.*, 2024], (3) *a few-shot*

[4]https://huggingface.co/intfloat/e5-base

[5]https://huggingface.co/Alibaba-NLP/gte-multilingual-reranker-base

Model	Methods	EVS	GSS	CGSS	ISD	LAPOP	Africa	Avg.
GPT-4o mini	Zero-shot Inference	0.5566	0.6026	0.4019	0.6109	0.4195	0.3923	0.4973
	Role-Assignment	0.5738	0.7564	0.4813	0.6164	0.4742	0.5563	0.5764
	Few-Shot Learning	0.5271	0.6538	0.4631	0.5804	0.4220	0.4258	0.5120
	Hybrid Method	0.5938	0.7292	0.5048	0.6330	0.4414	0.5305	0.5721
	ValuesRAG†	**0.6020***	**0.7781***	**0.5387***	**0.7001***	**0.5030***	**0.5953***	**0.6195***
Gemini 1.5 Flash	Zero-shot Inference	0.5419	0.6408	0.4502	0.6017	0.4149	0.4181	0.5113
	Role-Assignment	0.5598	0.7493	0.4770	0.6048	**0.4747**	0.5262	0.5653
	Few-Shot Learning	0.5225	0.6376	0.4559	0.5782	0.4194	0.4758	0.5149
	Hybrid Method	0.5845	0.7193	0.5026	0.6253	0.4448	0.5166	0.5655
	ValuesRAG†	**0.5869**	**0.7686***	**0.5337***	**0.6789***	0.4705	**0.5473***	**0.5977***

Table 2: **Accuracy scores for various methods compared with multiple baselines across six regional datasets.** k indicates the number of summaries to be retrieved. **Bold text** indicates the best performance, underlined text indicates the second-best performance. * denotes significant improvements (paired t-test with Holm-Bonferroni correction, $p < 0.05$) over all baseline model(s). † denotes our proposed method.

Model	Num(K)	EVS	GSS	CGSS	ISD	LAPOP	Africa	Avg.
GPT-4o mini	1	0.5960	0.7722	0.5347	0.6853	0.4682	0.5905	0.6078
	3	**0.6021**	**0.7781**	**0.5387**	0.7001	0.4686	**0.5953**	0.6138
	5	0.6052	0.7706	0.5301	**0.7016**	**0.5061**	0.5905	**0.6174**
	10	0.6020	0.7380	0.5317	0.7014	0.5030	0.5680	0.6074
Gemini 1.5 Flash	1	0.5753	0.7668	0.5272	0.6646	0.4548	0.5369	0.5876
	3	**0.5869**	0.7686	**0.5337**	**0.6789**	**0.4705**	0.5473	**0.5977**
	5	0.5868	**0.7690**	0.5303	0.6734	0.4661	**0.5498**	0.5959
	10	0.5852	0.7665	0.5279	0.6773	0.4509	0.5464	0.5924

Table 3: **Accuracy scores across six regional datasets for the ablation study (Section 5.3.2)**. Num(K) ($k \in \{1, 3, 5, 10\}$) indicates the number of demographic summaries retrieved. **Bold text** indicates the best performance, underlined text indicates the second-best performance.

learning method [Choenni and Shutova, 2024], and (4) *a hybrid method that combines both (1) and (2)*.

Specifically, for the role-assignment baseline, we use the same demographic summaries as in ValuesRAG to ensure fairness by assigning roles based on demographic information from the survey data. For the few-shot method, we follow the approach outlined in the previous work, where we randomly select five examples from the test set as prompts. The hybrid method combines both strategies, assigning roles based on demographic summaries and augmenting the prompts with five randomly selected few-shot examples from the test set.[6]

Additionally, We use *ValuesRAG* with $k = 3$ retrieved summaries—chosen to provide a good balance between retrieval diversity and contextual relevance.[7]

Evaluation. We utilize accuracy as the primary evaluation metric by converting multiple-choice responses into a binary format for consistency and simplicity. This is achieved by categorizing responses into two distinct groups based on contrasting answer patterns, effectively capturing agreement and disagreement. This approach aligns with the structure of values-related questions, such as those using a Likert Scale.

[6]Detailed prompts are provided in Appendix B.

[7]Detailed analysis of varying k and its implications on model performance is provided in Section 5.3.1.

5.2 Experimental Analysis

In our experiments, we compare our method, *ValuesRAG*, with four baseline methods: zero-shot inference, role-assignment, few-shot learning, and a hybrid approach that combines role-assignment and few-shot learning, as shown in Table 2.

We find that the role-assignment method generally surpasses both zero-shot and few-shot approaches. By grounding the agent's responses in a clearly defined demographic context, it ensures more consistent performance. Yet, role assignment can sometimes lead to overly narrow representations when demograhpic roles are interpreted stereotypically. Meanwhile, few-shot learning can incorporate example-driven context, but its limited number of prompts may not consistently address the intricate ways individuals' belifs diverge within similar social settings. As a result, it struggles to generalize to the multifaceted nature of human values, particularly when faced with unexpected or complex cultural scenarios. The hybrid method, which merges role assignment and few-shot prompts, does offer a partial improvement in contextual diversity. However, it remains insufficient for capturing the full spectrum of nuances that can arise from overlapping demographic factors and idiosyncratic personal persepectives.

In contrast, ValuesRAG overcomes these challenges by dynamically retrieving and integrating specific values-related cultural data for each agent. By focusing on values as the primary retrieval targets, this retrieval-augmented framework enables the model to include an expansive set of contextual clues, helping it reflect the depth and breadth of each individual's background and values. Crucially, our ValuesRAG provides a more adaptive mechanism for representing the subtle interplay of personal beliefs and cultural norms via avoiding the limitations of rigid demographic labels or small-sample prompts. ValuesRAG more effectively captures the complex dynamics that can shape a respondent's stance on different questions. Evaluations across diverse test datasets demonstrate that ValuesRAG with $k = 3$ consistently outperforms baseline methods, highlighting its ability to better represent cultural diversity, improve contextual alignment, and enhance overall model performance.

5.3 Ablation Study

We conduct two ablation studies to analyze the configuration and robustness of ValuesRAG., in Section 5.3.1, we vary the number of retrieved summaries (k) to examine how retrieval depth affects the model's performance. Second, in Section 5.3.2, we isolate the effect of using only values-based generation.

5.3.1 Varying the Number of Retrieved Summaries

To further investigate the role of retrieval depth in *ValuesRAG*, we vary the number of retrieved summaries, $k \in \{1, 3, 5, 10\}$. Table 3 presents the performance of *ValuesRAG* for three models across multiple datasets. We observe that while $k = 1$ may not provide enough diversity, increasing k beyond 3 can sometimes introduce less relevant information.

Moreover, $k = 3$ consistently achieves significantly better accuracy compared to higher values of k, highlighting its effectiveness and achieving the best balance between contextual relevance and model latency; these ensure both effectiveness and efficiency. This configuration not only enhances retrieval accuracy but also maintains low computational overhead, underscoring the strength of *ValuesRAG* in real-world applications where speed and performance are equally crucial. Hence, we choose $k = 3$ as the default retrieval number for conducting our baseline experiments, ensuring optimal performance across various models and datasets.

5.3.2 Impact of Values-Only Generation

To validate the robustness of ValuesRAG, we perform an ablation study using only values context augmented generation, thereby excluding the impact of demographic summaries on the model's performance. We use the WVS validation set—separated from the training data, which served as the retrieval corpus (as outlined in Section 3)—to evaluate the models. Table 4 presents a comparison of our method, using *only summaries of values*, against four baseline methods. Notably, ValuesRAG consistently outperforms all baselines across this validation data, achieving the highest accuracy despite relying exclusively on the values summaries.

This result confirms the effectiveness and robustness of the values-augmented generation approach. ValuesRAG leverages structured values summaries to generate contextually rich and culturally aligned responses. Even without demographic augmentation, ValuesRAG achieves superior performance by dynamically capturing the underlying value patterns, demonstrating its ability to generalize across diverse cultural contexts without requiring predefined QA examples or demographic anchors. The results demonstrate the framework's scalability and adaptability, effectively mitigating biases and generating culturally coherent outputs with minimal dependence on external context.

Methods	**GPT-4o-mini**	**Gemini-1.5-flash**
Zero-Shot	0.6176	0.6041
Role-Assignment	<u>0.6747</u>	<u>0.6505</u>
Few-Shot Learning	0.6359	0.6086
Hybrid Method	0.6670	0.6354
Values Augmented	**0.6894**	**0.6583**

Table 4: Accuracy comparison between baseline methods and our proposed Values Augmented Generation method using the WVS validation set. **Bold text** indicates the best performance, <u>underlined text</u> indicates the second-best performance.

6 Conclusion

We present *ValuesRAG*, an innovative Retrieval-Augmented Generation framework that revolutionizes cultural values alignment through its comprehensive approach to values context analysis. While existing approaches rely on pretrained knowledge with fixed demographic labels, often leading to oversimplified generalizations and biases, or few-shot prompting techniques that struggle with generalization due to limited correlations across value dimensions, ValuesRAG overcomes these weaknesses by dynamically retrieving and integrating granular demographic data with detailed individual-level cultural profiles. Our extensive evaluations demonstrate that ValuesRAG achieves an accuracy improvement of up to 21% over state-of-the-art baselines. By combining adaptive retrieval mechanisms with in-context learning and reranking strategies, *ValuesRAG* captures complex cultural nuances, reduces biases, and ensures contextually aligned responses. This structured design enables *ValuesRAG* to bridge the gap between generic LLM capabilities and the demands of culturally sensitive applications, providing a scalable and robust solution for real-world use cases.

7 Limitation

Although our baseline comparisons indicate that ValuesRAG generally delivers superior performance compared to alternative methods, it does not always guarantee an exact match to individual's true values. Since we rely on the WVS dataset to summarize individual profiles, there can be mismatches when these summaries are applied to other test sets. In future work, we plan to explore more adaptive retrieval strategies that can better align with novel datasets, as well as investigate how integrating additional fine-tuning with retrieval-augmented generation may further refine each agent's contextual accuracy.

References

[Achiam *et al.*, 2023] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[Afrobarometer, 2023] Afrobarometer. Afrobarometer data, merged round 9 data (39 countries), 2023, 2023.

[Bian and Li, 2012] Yanjie Bian and Lulu Li. The chinese general social survey (2003-8). *Chinese Sociological Review*, 45(1):70–97, 2012.

[Caliskan *et al.*, 2017] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, April 2017.

[Choenni and Shutova, 2024] Rochelle Choenni and Ekaterina Shutova. Self-Alignment: Improving Alignment of Cultural Values in LLMs via In-Context Learning, August 2024.

[Davern *et al.*, 2024] Michael Davern, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. General social survey 1972-2024. Data accessed from the GSS Data Explorer website, 2024. Machine-readable data file.

[Durmus *et al.*, 2024] Esin Durmus, Karina Nguyen, Thomas I. Liao, Nicholas Schiefer, Amanda Askell, Anton Bakhtin, Carol Chen, Zac Hatfield-Dodds, Danny Hernandez, Nicholas Joseph, Liane Lovitt, Sam McCandlish, Orowa Sikder, Alex Tamkin, Janel Thamkul, Jared Kaplan, Jack Clark, and Deep Ganguli. Towards Measuring the Representation of Subjective Global Opinions in Language Models, April 2024.

[EVS, 2022] EVS. European Values Study 2017: Integrated Dataset (EVS 2017)European Values Study 2017: Integrated Dataset (EVS 2017), 2022.

[Gallegos *et al.*, 2024] Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics*, 50(3):1097–1179, September 2024.

[GLM *et al.*, 2024] Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, et al. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*, 2024.

[Haerpfer *et al.*, 2022] Christian Haerpfer, Ronald Inglehart, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Jaime Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, and Bi Puranen. World values survey wave 7 (2017-2022) cross-national data-set. *(No Title)*, 2022.

[Hu *et al.*, 2024] Tiancheng Hu, Yara Kyrychenko, Steve Rathje, Nigel Collier, Sander van der Linden, and Jon Roozenbeek. Generative language models exhibit social identity biases. *Nature Computational Science*, pages 1–11, 2024.

[Huang *et al.*, 2024] Linus Ta-Lun Huang, Gleb Papyshev, and James K. Wong. Democratizing value alignment: From authoritarian to democratic AI ethics. *AI and Ethics*, pages 1–8, December 2024.

[Ji *et al.*, 2024] Jiaming Ji, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Kwan Yee Ng, Juntao Dai, Xuehai Pan, Aidan O'Gara, Yingshan Lei, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, and Wen Gao. AI Alignment: A Comprehensive Survey, May 2024.

[Jiang *et al.*, 2023] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023.

[Karinshak *et al.*, 2024] Elise Karinshak, Amanda Hu, Kewen Kong, Vishwanatha Rao, Jingren Wang, Jindong Wang, and Yi Zeng. LLM-GLOBE: A Benchmark Evaluating the Cultural Values Embedded in LLM Output, November 2024.

[Kharchenko *et al.*, 2024] Julia Kharchenko, Tanya Roosta, Aman Chadha, and Chirag Shah. How Well Do LLMs Represent Values Across Cultures? Empirical Analysis of LLM Responses Based on Hofstede Cultural Dimensions, June 2024.

[Lab, 2021] LAPOP Lab. Americasbarometer dominican republic, 2021, v1.2, 2021. Accessed on [Add the access date if needed].

[Li *et al.*, 2024] Hao Li, Ruoyuan Gong, and Hao Jiang. Political actor agent: Simulating legislative system for roll call votes prediction with large language models. *arXiv preprint arXiv:2412.07144*, 2024.

[Liu *et al.*, 2024a] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

[Liu *et al.*, 2024b] Yang Liu, Yuanshun Yao, Jean-Francois Ton, Xiaoying Zhang, Ruocheng Guo, Hao Cheng, Yegor Klochkov, Muhammad Faaiz Taufiq, and Hang Li. Trustworthy LLMs: A Survey and Guideline for Evaluating Large Language Models' Alignment, March 2024.

[Manvi *et al.*, 2024] Rohin Manvi, Samar Khanna, Marshall Burke, David Lobell, and Stefano Ermon. Large Language Models are Geographically Biased, October 2024.

[Masoud *et al.*, 2024] Reem I Masoud, Martin Ferianc, Philip Colin Treleaven, and Miguel RD Rodrigues. Llm alignment using soft prompt tuning: The case of cultural alignment. In *Workshop on Socially Responsible Language Modelling Research*, 2024.

[Potter *et al.*, 2024] Yujin Potter, Shiyang Lai, Junsol Kim, James Evans, and Dawn Song. Hidden Persuaders: LLMs' Political Leaning and Their Influence on Voters, November 2024.

[Sahgal and Evans, 2021] Neha Sahgal and Jonathan Evans. India survey dataset, 2021.

[Shen *et al.*, 2023] Tianhao Shen, Renren Jin, Yufei Huang, Chuang Liu, Weilong Dong, Zishan Guo, Xinwei Wu, Yan Liu, and Deyi Xiong. Large Language Model Alignment: A Survey, September 2023.

[Singh *et al.*, 2024] Shivalika Singh, Angelika Romanou, Clémentine Fourrier, David I. Adelani, Jian Gang Ngui, Daniel Vila-Suero, Peerat Limkonchotiwat, Kelly Marchisio, Wei Qi Leong, Yosephine Susanto, Raymond Ng, Shayne Longpre, Wei-Yin Ko, Madeline Smith, Antoine Bosselut, Alice Oh, Andre F. T. Martins, Leshem Choshen, Daphne Ippolito, Enzo Ferrante, Marzieh Fadaee, Beyza Ermis, and Sara Hooker. Global MMLU: Understanding and Addressing Cultural and Linguistic Biases in Multilingual Evaluation, December 2024.

[Sun *et al.*, 2021] Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiaxiang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, et al. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint arXiv:2107.02137*, 2021.

[Tao *et al.*, 2024] Yan Tao, Olga Viberg, Ryan S Baker, and René F Kizilcec. Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9):pgae346, 09 2024.

[Team *et al.*, 2024] Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*, 2024.

[Touvron *et al.*, 2023] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[Wang *et al.*, 2022] Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*, 2022.

[Webster *et al.*, 2021] Kellie Webster, Xuezhi Wang, Ian Tenney, Alex Beutel, Emily Pitler, Ellie Pavlick, Jilin Chen, Ed Chi, and Slav Petrov. Measuring and Reducing Gendered Correlations in Pre-trained Models, March 2021.

[Xie *et al.*, 2024] Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Shiyang Lai, Kai Shu, Jindong Gu, Adel Bibi, Ziniu Hu, David Jurgens, James Evans, Philip Torr, Bernard Ghanem, and Guohao Li. Can Large Language Model Agents Simulate Human Trust Behavior?, November 2024.

[Yoo *et al.*, 2024] Kang Min Yoo, Jaegeun Han, Sookyo In, Heewon Jeon, Jisu Jeong, Jaewook Kang, Hyunwook Kim, Kyung-Min Kim, Munhyong Kim, Sungju Kim, et al. Hyperclova x technical report. *arXiv preprint arXiv:2404.01954*, 2024.

[Zhang *et al.*, 2024] Xin Zhang, Yanzhao Zhang, Dingkun Long, Wen Xie, Ziqi Dai, Jialong Tang, Huan Lin, Baosong Yang, Pengjun Xie, Fei Huang, Meishan Zhang, Wenjie Li, and Min Zhang. mgte: Generalized long-context text representation and reranking models for multilingual text retrieval, 2024.

A Appendix A. Dataset Description

A.1 WVS

As shown in Figure 3, the WVS dataset conducts surveys in over 120 countries worldwide, covering most regions and countries across the globe, making it suitable for constructing a documents corpus for Retrieval Augmented Generation (RAG). It is also representative in terms of demographic characteristics. We select the most recent WVS survey (Wave 7) to ensure the dataset is as close as possible to the contemporary world.

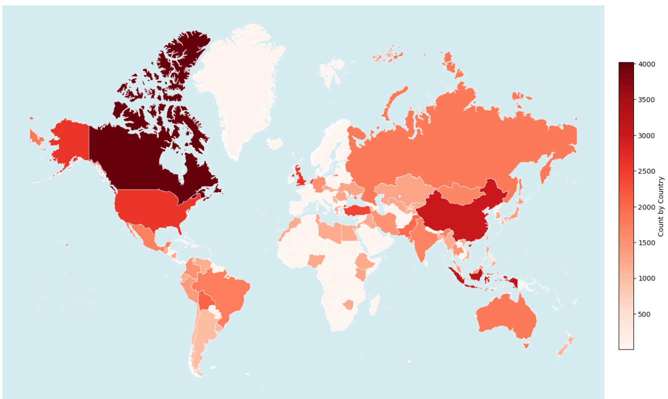

Figure 3: **Distribution of Country in WVS dataset.** The WVS dataset is the most extensive values survey dataset in terms of coverage.

According to the classification in the codebook released by WVS, the values-related questions it contains can be divided into 13 topics, as shown in Table 5, covering most dimensions of values, allowing for a comprehensive and accurate measure of each respondent's values.

Topic	**Count**
Social Values, Norms, Stereotypes	45
Happiness and Wellbeing	11
Social Capital, Trust and Organizational Membership	47
Economic Values	6
Perceptions of Corruption	9
Perceptions of Migration	10
Perceptions of Security	21
Index of Postmaterialism	6
Perceptions about Science and Technology	6
Religious Values	12
Ethical Values	23
Political Interest and Political Participation	35
Political Culture and Political Regimes	25

Table 5: **Distribution of Values-related Questions in WVS.** The questions were categorized into 13 topics with a total of 259 questions covering most of the dimensions of values

A.2 Test Datasets

To validate the robustness of our method, we selected datasets from six different regions as test sets. These test sets include both values-related questions and demographic characteristics. The demographic characteristics are used to generate summaries, serving as retrieval targets for RAG. The values-related questions are utilized as test questions to calculate accuracy (ACC).

EVS The first dataset comes from the *European Values Study*[8], a large-scale, cross-national, and longitudinal survey research program designed to explore values, beliefs, and attitudes across Europe. This dataset includes a total of 211 values-related questions and captures 34 demographic characteristics of the respondents. We select EVS 2017, conducted in 2017, ensuring alignment with the World Values Survey (WVS) in terms of the timeframe.

GSS We select GSS to represent the population of the United States. *The General Social Survey*[9] is a sociological survey that has been conducted since 1972 by the National Opinion Research Center (NORC) at the University of Chicago. Its primary purpose is to collect and analyze data on the opinions, behaviors, and demographic characteristics of adults in the United States, thereby monitoring societal change and the growing complexity of American society. Its questionnaire covers a comprehensive and wide range of topics, including many values-related questions. Specifically, within the GSS, we identify 44 questions as values-related and 33 questions as demographic characteristics.

CGSS *The Chinese General Social Survey*[10], initiated in 2003, is China's earliest national, comprehensive, and continuous academic survey project. Conducted by the National Survey Research Center at Renmin University of China, the CGSS systematically collects data at multiple levels, including society, communities, families, and individuals. CGSS only provided the questionnaire and data in Chinese, which we have translated into English to ensure its usability. We ultimately compile a total of 58 values-related questions and 13 demographic characteristics.

ISD To ensure that our experiment covers as much of the world's population as possible, we made efforts to include India within the scope of our test set. However, we were unable to obtain data from several government surveys in India, so we used data published by the Pew Research Center instead. The Pew Research Center's *India Survey Dataset*[11] is a comprehensive resource that captures the perspectives of 29,999 Indian adults on various aspects of society, including religious beliefs and practices, identity, nationalism, and societal tolerance. Conducted through face-to-face interviews between November 17, 2019, and March 23, 2020, the survey encompassed participants from diverse religious backgrounds, such as Hindus, Muslims, Sikhs, Christians, Buddhists, Jains, and others. This dataset covers 33 values-related questions and 23 demographic characteristics.

[8] https://europeanvaluesstudy.eu

[9] https://gss.norc.org

[10] http://cgss.ruc.edu.cn

[11] https://www.pewresearch.org/dataset/india-survey-dataset

LAPOP The Latin American Public Opinion Project (LAPOP)[12] is a research institute based at Vanderbilt University in Nashville, Tennessee. LAPOP's most notable survey is the *AmericasBarometer*, the most extensive survey of democratic public opinion and behavior covering the Americas, including North, Central, South America, and the Caribbean. This survey measures democratic values and behaviors through voter surveys, providing valuable insights into public sentiments across the region. We select this dataset to represent the population of Latin America. There are 48 values-related questions and 12 demographic characteristics.

Africa *Afrobarometer*[13] is a pan-African, non-partisan research network established in 1999 that conducts public attitude surveys on democracy, governance, economic conditions, and related issues across Africa. We selected data from the 8th round of Afrobarometer (collected in 2022), covering 34 African countries. After screening, a total of 144 values-related questions and 14 demographic characteristics are obtained.

B Appendix B. Prompts Used

This section provides the prompts designed for various components of the ValuesRAG, including prompts for performing question answering tasks, as well as for generating values and demographic summaries.

Prompt for Question Answering

Task: Respond to the question as the target individual, selecting the answer that aligns with their values and demographic context.

Rules:

- Step-by-step analysis using retrieved demographics and values data.
- Maintain the target individual's perspective throughout the analysis.
- Provide the response in JSON format, without additional explanation.

Steps for Inferring:

1. Analyze the demographics (age, gender, cultural background, social class, religion, and economic class) of retrieved individuals. Compare them with the target individual.
2. Identify individuals whose demographics most closely match the target individual. Note their IDs.
3. Based on the matched individual's values, infer how the target would respond.
4. Select the response that best aligns with the inferred values, and return only the integer representing the selected option.

[12] https://www.vanderbilt.edu/lapop

[13] https://www.afrobarometer.org

Prompt for Values Summary Generation

You are a summarization expert with expertise in extracting key insights from complex data. Based on the provided context, summarize this person's values in one paragraph.

Prompt for Demographic Summary Generation

You are a summarization expert with expertise in extracting key insights from complex data. Based on the provided context, summarize this person's demographics in one paragraph.

C Appendix C. Case Study

Example 1

Question

Is premarital sex always wrong, almost always wrong, sometimes wrong, or not wrong at all?

Original Demographics Summary

The respondent is a 63-year-old male born in the United States, of White descent, who speaks English at home. He completed 12th grade and holds a high school diploma, while his spouse has the same educational background. They are currently married and live in a household of two, with no children. The respondent identifies as middle class and reported a family income of $90,000 to $109,999 last year, with his own earnings falling within the same range. He practices Protestantism and is affiliated with the United Presbyterian Church in the U.S.A. His mother has completed two years of college, while his father finished high school. The respondent's demographic background reflects a stable family life with no financial challenges reported.

Retrieved Summaries:

Retrieved Text 1

Demographics The respondent is a 60-year-old White, non-Hispanic male born in the United States, where he speaks English at home. He was raised in a family where both parents were also born in the U.S., and he holds a bachelor's degree, signaling a level of educational attainment typical of the upper middle class. He is currently employed full-time in a professional and technical role within the private business sector and is the chief wage earner in his household, which consists of one person. The respondent is divorced and has two children. He identifies as a Protestant and has reported saving money over the past year, rating his household income as relatively high on a scale of one to ten.

Values This person's values center on family, personal empowerment, and social equity, emphasizing a balanced life encompassing work, political engagement, and community well-being. They prioritize financial responsibility, health, and individual liberties while advocating equality and rejecting traditional gender roles. They approach trust cautiously, emphasize accountability in governance, and hold nuanced views on immigration and moral integrity. They believe in science and technology's benefits and maintain personal faith amid diverse beliefs. Politically active, they support democratic principles and civil rights while voicing dissatisfaction with current institutions. Overall, they blend progressive values focused on social justice, environmental sustainability, and the power of citizen engagement in politics and public life.

Retrieved Text 2

Demographics The respondent is a 63-year-old White, non-Hispanic male born in the United States, where he speaks English at home. He is a citizen of the U.S., with both parents also being born in the country. He is married and lives in a household of two people, including himself and his spouse, who is retired. He has two children and both he and his spouse completed 12th grade, earning a high school diploma. He works full-time in a semi-skilled occupation within the private sector and identifies as belonging to the middle class, reporting that his household income places them in the third income tier. The respondent aligns with the Protestant religion and describes their financial situation as stable.

Values This individual values family and work as life's center, while moderately engaging in politics. They emphasize independence, hard work, tolerance, and determination in children, embrace progressive stances on gender and social issues, and place little importance on religious conformity. Their trust is selective, relying heavily on family, showing moderate faith in others, and opposing corruption, while valuing economic equality and environmental concerns. They uphold science and technology, favor personal freedom over security, and maintain moral integrity and accountability. Strongly engaged in civic matters, they are committed to democracy, civil rights, and active political participation, yet remain cautious about government roles. Overall, their values balance familial dedication, personal agency, ethical living, and a progressive yet pragmatic outlook on social and political issues.

Retrieved Text 3

Demographics The respondent is a 65-year-old white male, born in the United States, ...

Example 2

Question

How many children to have is a personal matter, and the government should not interfere. Do you agree?

Original Demographics Summary

The respondent is a 27-year-old male from China of Han ethnicity, who speaks Chinese at home. He holds a postgraduate degree and has never engaged in any work for economic income. Currently, he lives alone without any additional household members and identifies as belonging to the middle class, describing his household's economic status as average in his local area. He does not practice any religion and reports no personal income in the past year.

Retrieved Summaries:

Retrieved Text 1

Demographics The respondent is a 28-year-old male of Chinese ethnicity, born in the country where he currently resides. He speaks Standard Chinese (Mandarin/Putonghua/Guoyu) at home and identifies as part of the working class. He has attained a Bachelor's degree and is employed full-time in a clerical role within the private sector. He lives in a household of four, including his own parents, and is currently married but has no children. Both he and his spouse have a similar educational background and are employed full-time. His family's financial situation indicates that they have spent some savings and borrowed money over the past year. He does not belong to any religious denomination and has reported his household income as being in the third income level out of ten.

Values This individual holds a complex set of values that place a significant emphasis on family, work, and personal responsibility, while also promoting income equality and valuing hard work as a means of success. They have a moderate stance on political engagement, expressing skepticism toward certain institutions while maintaining high trust in organizations like the police and government. Their views reflect a critical perspective on issues such as immigration and gender roles, where they value cultural diversity but prioritize native citizens for job opportunities. They exhibit a commitment to environmental protection, scientific advancements, and ethical living, while acknowledging their non-religious identity. Overall, their values convey a blend of personal autonomy, accountability, and a pragmatic view on governance, demonstrating a preference for democratic principles intertwined with the belief in the importance of social order and national security.

Retrieved Text 2

Demographics The respondent is a 29-year-old male of Chinese descent, and speaks other Chinese dialects at home. He has attained a bachelor's degree and works full-time in the sales sector within a private business or industry. Living alone, he is single and has no children. His family background includes parents with upper secondary education. Socially, he identifies with the working class and describes his household's financial status as just getting by, placing his household income in the fourth income group on a scale of ten. He does not identify with any religious denomination.

Values This person exhibits a set of values that emphasize the importance of family, hard work, and personal responsibility, reflecting a traditional view of societal roles and a commitment to instilling these qualities in children. They hold a pragmatic and somewhat secular worldview, expressing low interest in politics and religion while maintaining a positive attitude towards health, personal agency, and economic stability. They advocate for income equality and environmental sustainability, demonstrating concern for social equity and opposing corruption, while also valuing trust primarily within familial ties. Additionally, their perspective on immigrants is cautiously appreciative, and they prioritize freedom and economic progress, alongside a recognition of the complexities of modern ethical and moral issues. Overall, this individual's values highlight a balance between familial loyalty, economic and environmental responsibilities, personal freedom, and a discerning approach to broader societal structures.

Retrieved Text 3

Demographics The respondent is a 25-year-old male of Chinese ethnicity, who speaks Standard Chinese languages, ...